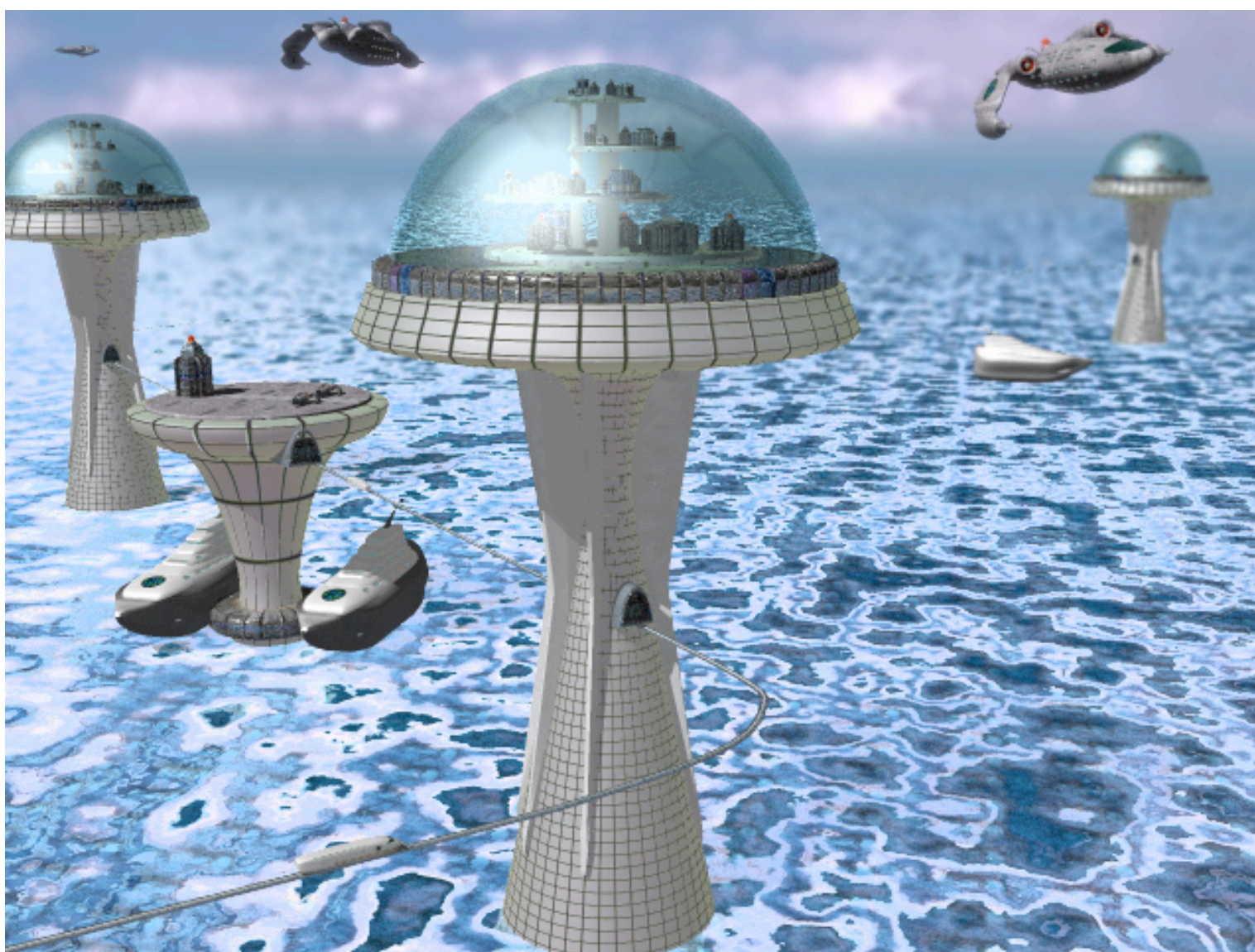


Beyond the OPAC: The Semantic Library

John Blyberg

Discovery & Access: The OPAC and beyond
Michigan Library Consortium
March 2, 2007

Looking to the Future



The description of today's event says that it will focus on "today's innovations and tomorrow's implications of the library OPAC."

I was originally going to address the "today's innovations" side of that equation. That was when I was still working for the Ann Arbor District Library. But now that I'm no longer working for them, I feel a little uneasy about devoting an entire talk to them.

Suffice to say that I feel very proud of the work I did on AADL's social OPAC. If you haven't seen it, I'd urge you to go to www.aadl.org and poke around. It certainly represents a new direction for the library OPAC.

But I've moved on from AADL, and so I'm going to shift my sights a little further ahead here as well and try to do a little soothsaying.

As you can imagine, predicting the future can be a little dicey.

No one knows, for sure, what the future will hold. I "searched" for "future" on Google images, and found this graphic--and this looks like a really fun place to visit, it's got a monorail snaking through

What is this?



Does this image look familiar to anyone?

Why would it? It's simply an industrial complex along-side a river, and obviously, it's not all built yet--though it will be.

Let me give you some hints.

This facility is built in a remote region of France

It's estimated cost is about 12.1 billion dollars

It serves no "practical" purpose

It's a Joint venture between the USA, the European Union, Japan, Korea, Russia, China, and now India.

It's the second-most expensive scientific project after the International Space Station

It will be decommissioned sometime around 2040

It's known as "ITER" – International Thermonuclear Experimental Reactor, and it's ultimate scientific goal is to produce 500 million watts of fusion power for 500 seconds.

But the REAL goal here is to find a way to generate a source of

What are we doing?



My point here, is that, as I look to the future, I see the possibility of major crisis within the library industry. And yes, it has everything to do with technology and the fact that we are constantly “chasing” it instead of driving it. There may come a point when we are just too far behind the curve--I don’t want to see that happen.

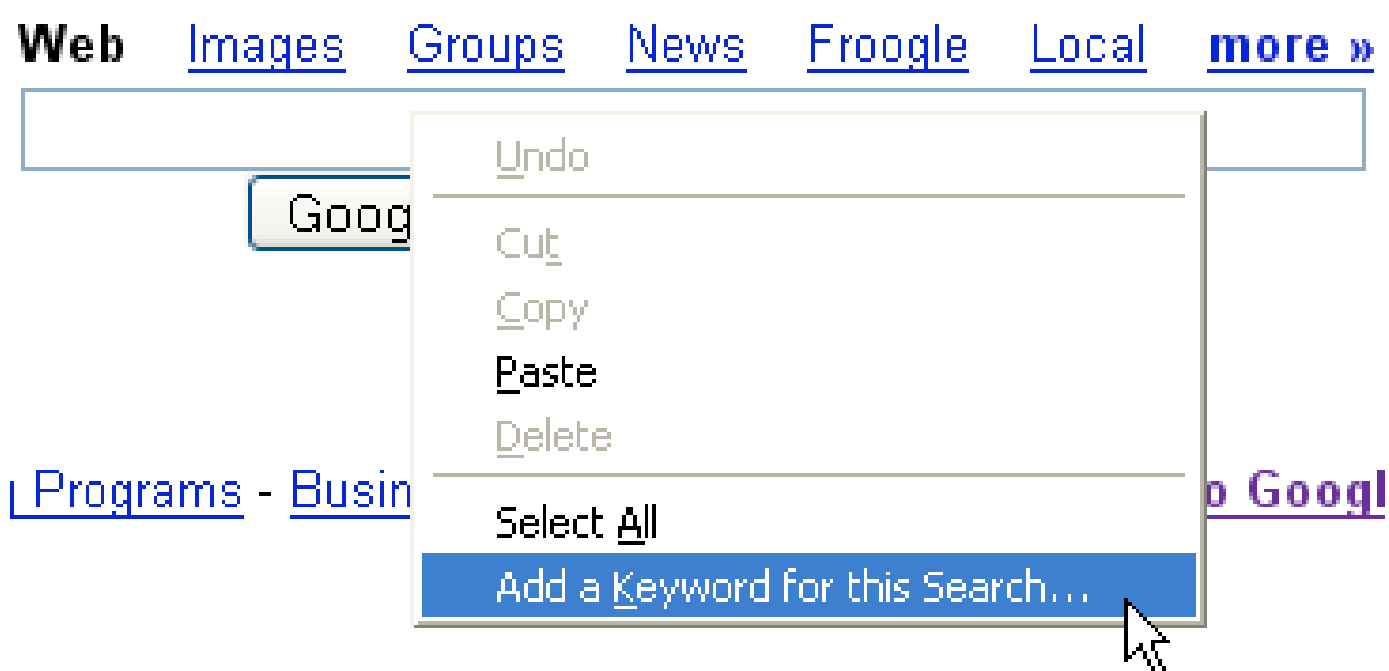
Last Fall, the “Lawrence Journal World” in Kansas ran an Op-Ed piece by a prominent local resident discouraging the construction of a new library in Lawrence on the premise that libraries are (quote) “Inefficient, Limited in their ability to provide information, and Obsolete”.

What makes this type of editorial disturbing is that more and more people are beginning to feel this way, and the reason for is that, even though his conclusions were way off base, there was some truth to his arguments.

I just wanted to mention this as an example, I don’t intend to debate its merits one way or the other, except to say that, in almost all cases, libraries are playing catch-up to technology--

“Thanks be to the Keyword”

Google™



That's because despite all of these converging values, and the inherent worth of social networking, we still find ourselves sitting in front of a search engine, wondering which keywords to use, and how to put them together.

The interface may change, and the software engine behind it, but the concept has remained the same for as long as we've been searching for "stuff" on computers. ("stuff" is the technical term for "data", by the way)

Incidentally, what do our users do when they can't find what they're looking for with a keyword?

They come see you!

Going beyond the keyword



What's the problem?

And, why do people go to librarians, and what makes librarians so valuable? What do I say to that person, in Lawrence Kansas when he asks why he needs a librarian?

Well, the keyword index is completely incapable of presenting relational information between concepts. The keyword search provides only “Matches” but cannot say, “oh, this reminds me of that, and you can find it here, and you know what else? This might be of help too!” That is called knowledge-based information retrieval and it's by-far the most effective research tool. It's the type of research strategy that allows us, as humans, to make the wild, connective leaps of thought that lead to major breakthroughs--it brought Newton to Gravity and Copernicus closer to the stars.

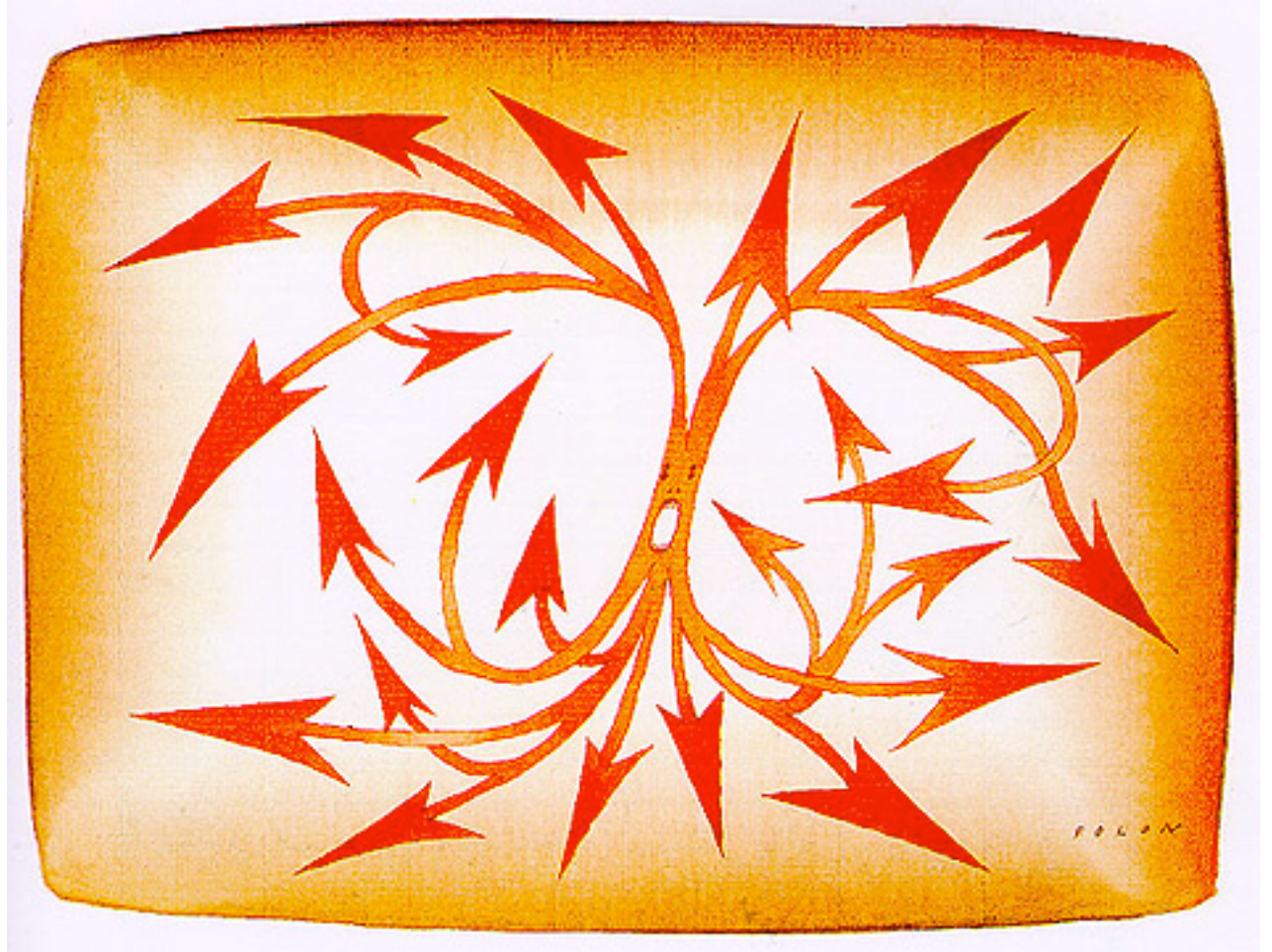
But there is a major, inherent limitation imposed upon knowledge-based information retrieval and that is that human brains are, for all intents and purposes, isolated networks.

For instance, you could bring the ten best oncologists together with the best medical library, and the best lab with an unlimited

The Semantic Web



- i. Web 1.0
- ii. Web 2.0
- iii. Semantic Web
(Web 3.0?)



Enter the Semantic Web.

The “Semantic Web” is a project being led by Tim Berners-Lee--the original creator of the web. As such, it’s got the backing of the World Wide Web Consortium--meaning it’s components will be open standards.

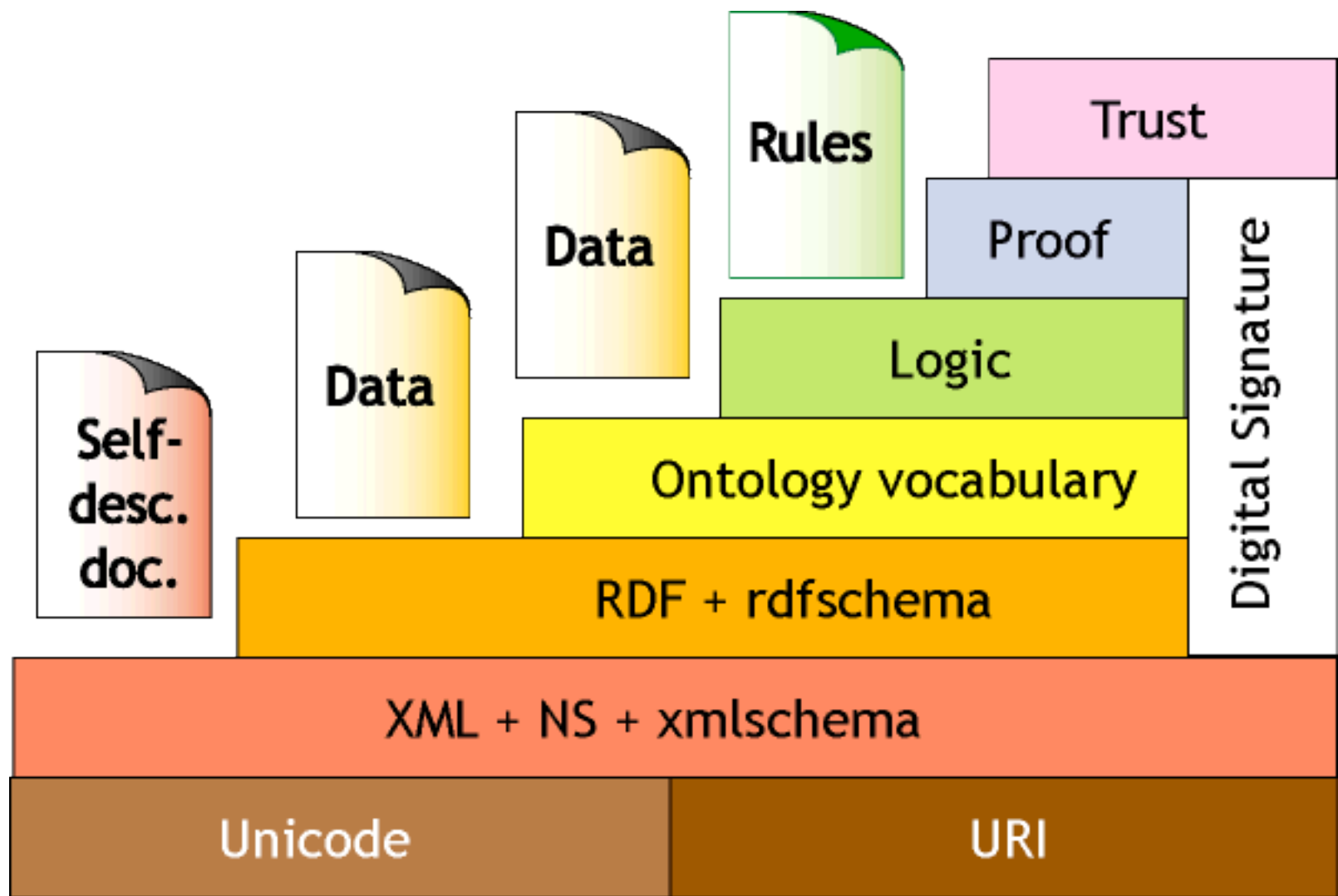
But what is it?

Their site says, “The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries”.

Sounds good, but it’s also a pretty vague description.

Presently, the web is a collection of documents--mostly written in HTML--even those pages that are dynamically generated from databases or user interaction are essentially static documents once rendered by the web browser. That’s why you can do a “save-this-page-as”.

The Rise of the Markup Language



The Semantic Web gathers a number of descriptive technologies to address the shortcoming of HTML: XML, RDF, and OWL (in that order as you travel up the stack here)

XML (extensible markup language) has been around for quite some time and has established itself as a very effective way of transmitting structured documents. It's important, however, to bear in mind that XML is a standard for syntactical format only, and does not impose any semantic constraints or meaning on the data it carries.

RDF (Resource Descriptive Framework) is the next step up this stack. It is a simple data model for carrying (and referring to) a payload of object data. In other words, a logical grouping of data. For example, an object could be a checked-out item, including all the necessary bit of information like, Title, Author, Duedate, etc. RDF is often represented in XML format. Can anyone think of a very popular RDF format? RSS is indeed an RDF object.

Both XML and RDF require a predetermined schema to describe their properties (and in RDF's case, it's objects and classes).

Inference

- User Agents

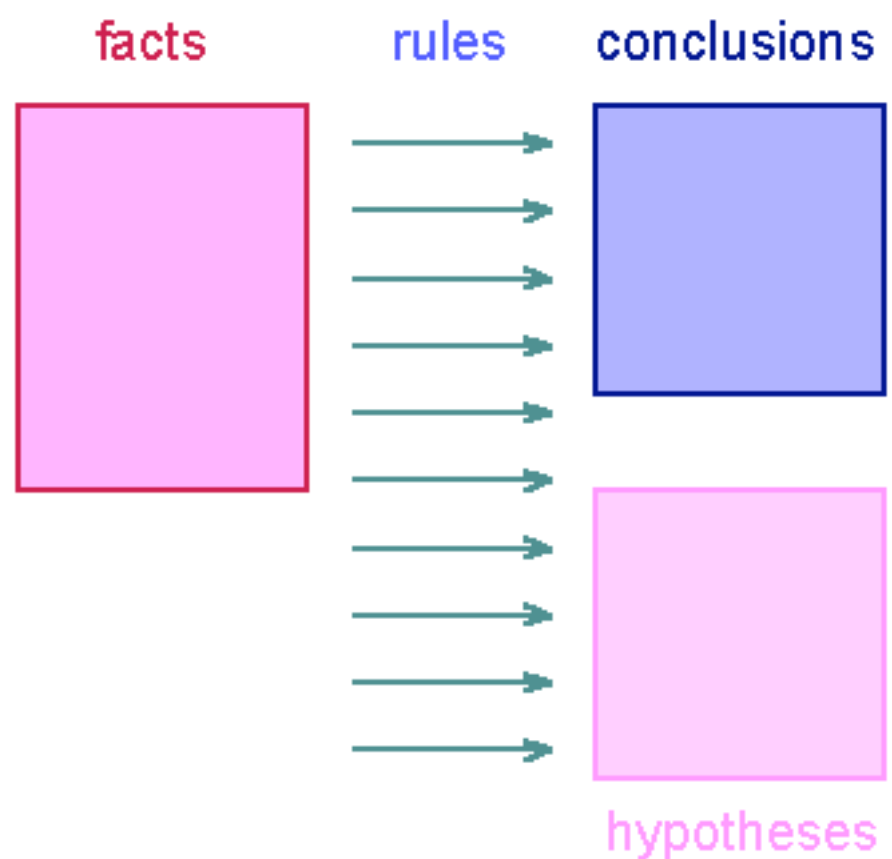
- Web Browser

- Human Interpretation

- Information Agents

- Consume Markup Language

- Processed by Inference Engines



Like I mentioned before, simple HTML is rendered by web browsers without any thought given to the content itself. That's because human eyeballs scan the data, and our human brain processes the information.

But the goal of the semantic web, is to make that same information available on a programmatic level so that it can be consumed and processed by software.

Remember, we want to be able to aid our knowledge-based information retrieval, and we can only do that, if somehow, our search engines can make intelligent, connective associations, draw conclusions and form hypotheses.

So the way Tim Berners-Lee imagines this working is that when semantic queries are run on a particular domain (subject area), a piece of software would scour the web for related ontologies, then process them and intelligently do all these things.

Imagine, for a minute, the example I used with the oncologists. Except, take every bit of cancer research ever done, in any

You will be assimilated!

(Not really)

The Semantic
Web is
“Weak AI”



Well, it is!

There's always someone who hears “artificial intelligence” and lets their mind fly off the handle into Borg-land.

If that's you, not to worry.

The Semantic Web is a type of “Weak Artificial Intelligence” That means that the system can not comprehend human-defined concepts. Instead, it indicates a machine's ability to solve a well-defined problem by performing a well-defined operation on well-defined data. In other word, the Semantic Web is a large-scale implementation of an “Expert System”.

And another term for “Expert Systems” is “Knowledge Based Systems”

Knowledge-based system--meet knowledge based information retrieval--meet librarians.

So these information agents will contain inference engines that are

Potential Benefits of SW

“I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A ‘Semantic Web’, which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The ‘intelligent agents’ people have touted for ages will finally materialize.”

Tim Berners-Lee

We can use the web to do a number of things we couldn't before. We can look for the closest pizza joint, we can reserve books at the library, we can participate in online auctions, we can go to sites and find the lowest price for a 40-inch plasma TV. We can do this because we can interpret the information presented to us on the screen.

A computer cannot do this for us, without direction, however. By presenting the web to humans and computers alike, in a standardized way, the Semantic Web opens the door to a new era of online computing.

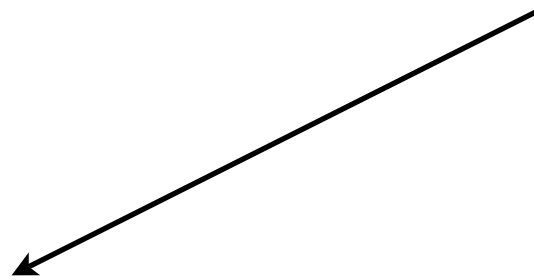
It's the stuff that we've read in science fiction novels--like electronic assistants that know it's time for you to get a haircut and automatically make that appointment and put it in your calendar. (I obviously need that)

But the people who are going to really benefit from this technology are the researchers--specifically, the scientific research community.

Semantic Publishing



Information
published online
as data objects



Semantic Markup/
Ontologies

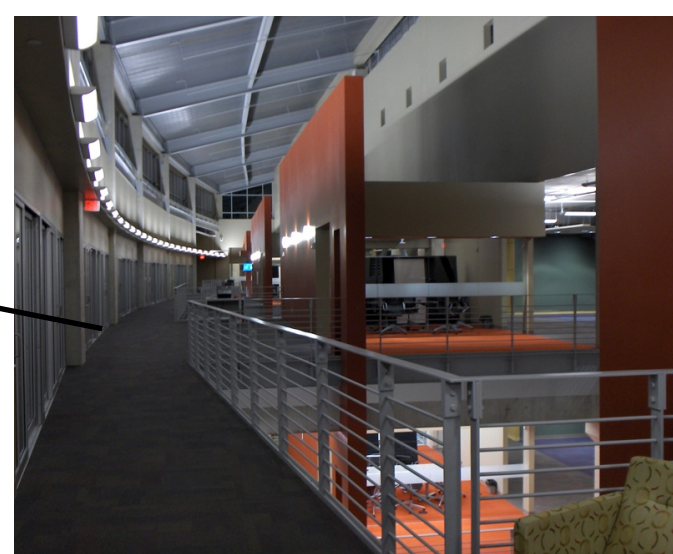


Semantic Publishing is expected to revolutionize Web and Scientific publishing.

One simple idea that may radically change scientific communication is for researchers to directly self-publish their experiment data in semantic format on the web. In one scenario, a scientist could design and run an experiment, and share the results with the world in real time by publishing the data as a semantic object on the web. Semantic search engines will make this semantic data available to everyone as it becomes available.

I also see a number of uses in the humanities as well, as ideas and theories are shared freely within academia.

The Semantic Library



Semantic
Language



For me, the possibilities of the Semantic Web invariably leads back to library service. The reason for this is that, ultimately, the best possible outcome for the Semantic Web involves librarians. Why?

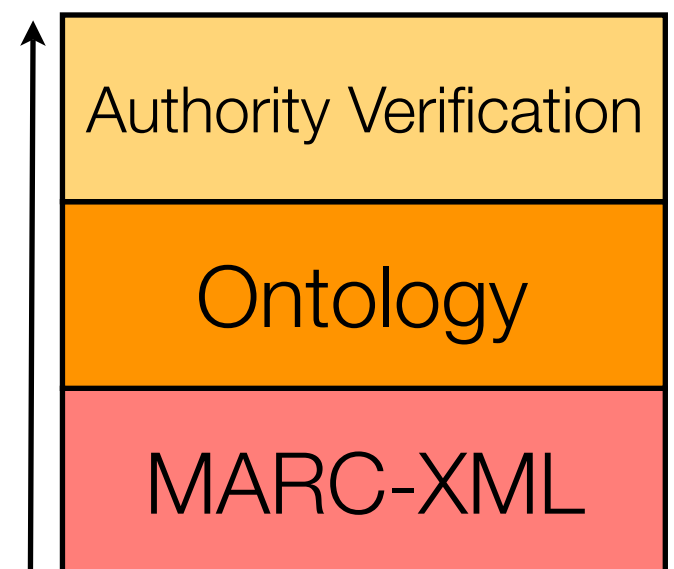
First, Libraries have the infrastructure in place already to take advantage of this technology. Librarians have the experience and knowledge necessary to piece together the appropriate metadata from our existing collections. From a computer's point of view, what is a subject heading, if not metadata. In addition, we've got the people necessary to add his additional layer of functionality to our systems. We are really the ones that are going to make this all come together in a cohesive way.

Second, Libraries are positioned already as information gatekeepers. That is our role in society--to bring information and people together. We can look upon the Semantic Web as an opportunity to develop a remarkable tool for humankind.

Third, we're not encumbered by the need to protect proprietary information--sharing the wealth of knowledge is in our bones--We're conditioned to share already. Our mission flies in the face

What is a Semantic Library?

- The Semantic Catalog
 - Ontology-rich catalogs
 - MARC-based schema?
 - Centralized information agents provide a cooperative approach to data sharing
- ILSs become semantically intelligent
 - Targeted searching/discovering



I like to think in terms of examples, so imagine if your family was planning a trip to France and each member is doing some travel research in order to put together an itinerary. It's possible that a Semantic Library system would be able to build a profile of each person and understand that they are searching for travel-related information (I know that "profile" is a dirty word, but get over it). So, for example, if one person has checked-out, or even searched for, material about Matisse, the semantic search result would yield different results than another member of the family who has recently checked out "The history of Airbus."

And because the larger semantic web, the library website could immediately suggest destinations, or offer to make reservations at certain restaurants, or for shows, tours, etc. In addition, it could also intelligently recommend material for check-out that could be brought along on the trip.

This may sound a little far-fetched, but sites like Amazon and Expedia do that now. The major difference is that while those sites are indeed expert systems, they are virtually sealed, including only their partners and vendors. The semantic web opens the web up

Possible Pitfalls

- Privacy
 - Censorship
 - Opting in?
- “Double Entry” / Increased Workloads
- Network-centric
- Cost?

Like anything, the semantic web is not without its detractors and critics that say it'll never work. They may be right. It may not work at all (highly unlikely outcome), or It may just not work as presented currently by the W3C.

There are a number of privacy concerns. The biggest, of which, is that the Semantic Web would make it much easier for any entity, government, or otherwise, to track and control the flow of information they do not want being disseminated. In particular, regimes like China who impose the “Great Firewall of China” would find it a lot easier to track dissidents. Critics fear that this would make automatic blocking of sites and servers very easy and much more prevalent.

Also, given the amount of personal information online, the semantic web will be able to much more intelligently profile (there's that dirty word again) all of its users. Some people are very uncomfortable with that.

Another criticism of the Semantic Web is that it would be much more time-consuming to create and publish content as there would need to be

Possible Threats

- Vendors
 - Unable/unwilling to make necessary changes
- Fear of technology
 - “Is this going to replace me?”
- Getting “buy-in”
- Inability to coordinate
- Apathy & indifference

There are other possible threats to the adoption of this technology that are specific to libraries.

And it's no coincidence that vendors are at the top of that list. One of the major problems we face is the fact that our ILSs are becoming more and more antiquated as time goes on. I've written before about the need for vendors to begin adopting standards-based APIs (that's application programming interface), such as XML-based web service layers. Some do with varying levels of quality, most don't, and, of course, some say they do, but don't really.

Identifying strategies for dealing with vendors in the long term is critical to our adaptability to new technology.

Another internal problem many of us face is the fear of technology among staff. And perhaps fear is not always the right word. Several months ago, Darien Public Library in Connecticut (where I'm headed) hosted an open-spaces conference called “Library Camp”, this was the second such gathering of its kind.

Going forward...

- Commitment to technology
- Strategic planning
- Technical coursework @ library school
- Continuing education
- Communication and dialogue
- Pay attention to World Wide Web around you!

The semantic web represents a new direction for the web, with a slew of associated hopes for a better, leaner, more efficient information architecture. I think it describes an inevitability--the standard, as it's written today, may not turn out exactly as it is on paper, but something very much like it is going to emerge. That's not a guess, it's simply the natural progression of things. You cannot argue to me that technology remains stagnant--especially not technology like that of the web.

So, we can either chase it, or we can get out in front of it, and maybe--if we're good, put our name on it in some way.

So, we need to ensure that our commitment to new technology remains firm. I think, in most cases, we want to commit to new technology, we're just not sure how.

There are a lot of ways we can do this, and maybe I'm full of hot air when I talk about a "Semantic Library", but the point is, that we can take our best guess as to where the future is going to take us, and maybe we can ambush a great opportunity to do something truly new and groundbreaking.

Resources

- World Wide Web Consortium (www.w3c.org)
- Dieter Fensel, Wolfgang Wahlster, Henry Lieberman, James Hendler (2002-11-15). Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential. MIT Press. ISBN 0-262-06232-1.
- Grigoris Antoniou, Frank van Harmelen (2004-04-01). A Semantic Web Primer. The MIT Press. ISBN 0-262-01210-3.

There is a lot of information available online about the Semantic Web--You can Google it if you like and start reading.

If you want to cut right to the chase, I suggest you go directly to the W3C.

Wikipedia also has some nice primers.

Thank-you, bla bla bla